RESEARCH PAPER

ATTENTION AGGREGATION U-NET++ (AAUNET++) AND DIFFUSION BASED VARIATIONAL AUTOENCODER (DVAE) FOR EPILEPTIC SEIZURE DETECTION

Ramya.K[1], Dr. M.Kokilamani[2]
Research Scholar, Department of Computer Science, Kamalam College of Arts and Science, Anthiyur, Affiliated to Bharathiar University, Coimbatore, Tamil Nadu, India,
Email ID: ramya.msc0709@gmail.com
Assistant Professor, Department of Computer Science, Kamalam College of Arts and Science, Anthiyur, Affiliated to Bharathiar University, Coimbatore, Tamil Nadu, India,
Email ID: Kokilamanikcas1@gmail.com

ABSTRACT: To monitor brain electrical activity and identify unusual spikes or asymmetries that can point to a seizure, electroencephalograms (EEG), are mostly utilised in epileptic seizure detection. However, it takes a lot of time for experts to visually examine an EEG, and their diagnoses could even conflict. For EEG diagnoses, an automated computer-aided diagnosis is therefore required. The Multivariate Variational Mode Decomposition (MVMD) algorithm for EEG signal denoising is suggested in this study. Construction of multivariate oscillations is based on the assumption that the same frequency element must be present in all channels of the input EEG data. Recovery of signals that have inherent multivariate modulated oscillations from multivariate input EEG data, a variational optimisation problem is developed for signal denoising. Attention Aggregation U-Net++ (AAUNET++) model has been developed for feature extraction. The original multi-level skip connections of U-net++ are replaced in the feature extraction process by attention aggregation, which is introduced to capture global contextual information. Diffusion Based Variational Autoencoder (DVAE) is introduced for seizure detection in EEG signals. DVAE integrates the probabilistic latent representation capability of a Variational Autoencoder (VAE) with the progressive noise modeling of a Denoising Diffusion Probabilistic Model (DDPM) to achieve robust feature learning and enhanced noise suppression in EEG. Using the Children's Hospital Boston-Massachusetts Institute of Technology (CHB-MIT) scalp EEG database, the recommended method attains superior efficiency by precision (PRE), recall/sensitivity, F-measure (F-score), and accuracy (ACC).

INDEX TERMS: Epileptic seizure detection, scalp EEG, Correlation Stacked Denoising Autoencoder (CSDAE), Improved Empirical Wavelet Transform (IEWT), and Residual Attention U-Net (RAUNet) model.
How to cite this article: Ramya.K, Dr. M.Kokilamani. Attention Aggregation U-Net++ (Aaunet++) And Diffusion Based Variational Autoencoder (Dvae) For Epileptic Seizure Detection. Int J Drug Deliv Technol. 2026;16(77s):1774-1783. DOI: 10.25258/ijddt.16.77s.201.

1. INTRODUCTION

Both adults and children can be affected by epilepsy, a neurological condition. One of its differentiating properties is abrupt recurrent epileptic seizures [1]. A quick, transient interference with a cluster of brain cells' electrical activity is essentially what causes this seizure disease. Epileptic seizures are brought on by the brain's neural networks experiencing excessive electrical activity. Involuntary movements brought on by these seizures may affect a whole body or only a section of it (generalised). They can also cause mood swings, cognitive abnormalities, loss of consciousness, and disturbances in sensation (hearing, vision, and taste) [2]. The frequency of seizures varies greatly, from less than once annually to many times daily. Four to five times as numerous individuals with active epilepsy death as those without seizures. Nonetheless, the risk of death is reduced by efficient medical treatment that is tailored for each patient. By systematically monitoring both seizures and the response to treatment, mortality can be decreased.

The human brain generates an electric signals called EEG that helps track the brain's development in medical treatment and shows the functions of the brain [3]. Computed tomography (CT) scans, magnetic resonance imaging (MRI), functional MRI (fMRI), amplitude-integrated EEG (aEEG), and continuous EEG (cEEG), and other methods are frequently used to identify epilepsy [4]. The aforementioned methods for detecting seizure early and aid in the effective treatment of epilepsy and may lower the risk of early mortality. Practitioners will save a great deal of time and effort by examining and evaluating the EEG signal much more easily with an automatic seizure detection system. A large number of traditional techniques examined EEG data in the time-frequency domain for automated seizure identification [5].

For epileptic seizures, Computer-aided detection (CAD) systems leverage Artificial Intelligence (AI) to analyze EEG signals, significantly improving the precision and efficiency of diagnosing epilepsy than manual, time-consuming human review. These systems, which often include pre-processing, feature extraction, and classification [5]. Among several methods, Variational Mode Decomposition (VMD) adaptively decomposes EEG signals, allowing researchers to discard noise-prone baseline components and retain specific modes related to seizure activity [6]. In order to eliminate noise and artifacts and improve accuracy, it efficiently breaks down complicated, non-stationary signals into band-limited intrinsic mode functions (VMFs). To remove noise, a threshold is determined for these coefficients, and those that fall below the range are deleted, and the denoised signal is reconstructed using the remaining coefficients [7].

Attention Aggregation UNet (Attention A U-Net) models are specialized architectures that enhance standard U-Net models for feature extraction by integrating attention gates (AGs) into skip connections. This approach automatically highlights salient features and suppresses irrelevant, noisy responses [8]. The main obstacles for automated seizure detection systems include the absence of clinical EEG data, the extremely skewed class distribution in seizures vs. non-seizures, patient to patient variations, and the problem of catastrophic forgetting when training sequentially with multiple patients. The Deep learning (DL) approaches, along with other machine learning (ML)-based classifiers, have been studied in the context of seizure detection from the EEG data set. Precision and accuracy have been utilised to determine how effectively these techniques perform for multi-label EEG classification. Therefore, utilising a feature extraction, DL may enhance the EEG-based epileptic seizure detection [9–10]. Time series data's global dependency is difficult for Convolutional Neural Networks (NN) to capture, but they are effective at extracting local features. Long-term dependence in time series can be captured by sequential (NN) neural network models such as Transformer, Recurrent NN (RNN), and Long Short Term Memory (LSTM). By encoding inputs as distributions rather than fixed points, VAE are generative DL models that learn a smooth, probabilistic latent space [11].

In this paper, MVMD algorithm is developed for EEG signal denoising. Variational optimisation is designed for signal denoising in order to extract noisy signals with intrinsic multivariate modulated oscillations from multivariate input EEG data. Attention Aggregation U-Net++ (AAUNET++) has been developed for feature extraction. A Diffusion-Based Variational Autoencoder (DVAE) is introduced to enhance EEG-based epileptic seizure detection by combining the strengths of a VAE and a DDPM to achieve robust feature learning and enhanced noise suppression in EEG. CHB-MIT database the proposed method has obtained highest precision, recall/sensitivity, F-measure, and accuracy than other methods.

2. LITERATURE REVIEW

By breaking down EEG data, the effective EEG-based method for detecting epileptic seizures (ES) was investigated by Das et al. [12]. In particular, the decomposed EEG signals in terms of IMFs through EMD are analyzed to extract features like variance, fluctuation index, and Ellipse Area Second Order Difference Plot (SODP). There were two types of composite feature formation techniques namely 1D and 2D. The characteristics of the EEG channels were the basis for the development of the former. Based to the research, the CNN model was selected for the 2D feature representation because it is the most widely used DL architecture for 2D inputs, while CNN for 1D feature was used to recognize the ES. The dataset used by the researchers was obtained from EEG data of the ES patients from Prince Hospital Khulna (PHK), Bangladesh, along with a standard dataset of CHB-MIT. When compared to a 2D composite feature form, the suggested technique shows promise for ES detection.

A framework for unsupervised learning to identify epileptic seizures employing a 1Dcascaded convolutional autoencoder (1D-CasCAE) was presented by Timothy Aboyeji et al. [13]. This method initially divides EEG recordings into 5 s epochs from a subset of patients in the CHB-MIT datasets. The correlation coefficient and Shannon entropy are used to choose eight informative channels. The 1D-CasCAE automatically learns the characteristic patterns of interictal data using downsampling and upsampling algorithms (non-seizure) segments. The model's resilience is greatly increased by incorporating a sliding window and adaptive thresholding, it may differentiate interictal segments in large EEG recordings. The proposed 1D-CasCAE successfully recognizes typical EEG signal patterns and takes use of reconstruction defects to identify aberrations (ictal segments), according to simulated results. Neurologists' manual EEG signal assessment process may be substituted for the developed technique in clinical situations. Furthermore, by using flexible time provides a to identify seizures, the suggested automated system may adjust to each patient's SOP.

The EEG corpus of Temple University Hospital, Raghu et al. [14] apply CNN and transfer learning to categorise seven seizure types with non-seizure EEG. Absence, focal non-specific, generalized non-specific, tonic, tonic-clonic, simple partial, complicated partial, and non-seizures are among the several forms of epileptic seizures that have been discussed. Before being fed into CNN, a spectrogram stack was produced from the entire set 19-channel EEG time series. The best network for the suggested study was found using the pretrained networks. Conventional feature and clustering-based methods were surpassed by CNN-based methods. In order to treat patients with epilepsy prior to surgery, the EEG-based CNN model for seizure type classification is deployed.

Dang et al. [15] suggested a novel method using DL with a multi-frequency multilayer brain network to identify epilepsy based on EEG. A multilayer brain network was developed using the brain's multi-frequency features and multi-channel EEG inputs. The multilayer network structure effectively detects epilepsy by mapping the EEG data's channel-related, time, and frequencies data. A multilayer deep CNN (MDCNN) model for classification is then produced using a multilayer brain network as input. The multilayer structure of the suggested brain network is precisely matched by the MDCNN model, this has a parallel multi-branch design in the initial section and is composed of two sections. It is designed to use multi-channel EEG signals to characterise complicated brain states.

Sahani et al. [16] multichannel scalp and single-channel EEG data are utilized to identify epileptic seizures using a Reduced Deep Convolutional Stack Autoencoder (RDCSAE) and an Improved Kernel Random Vector Functional Link Network (IKRVFLN). In order to train the suggested IKRVFLN classifier effectively by lowering the error for identifying promising epileptic seizure activity, the best discriminative unsupervised features are extracted from EEG data using the RDCSAE framework. The suggested approach is tested utilizing Boon University's single-channel EEG datasets and Boston Children's Hospital's benchmark multichannel scalp EEG (sEEG). The suggested approach is used to build a CAD system on a fast, hardware environment for reconfigurable Field-Programmable Gate Array (FPGA).

Liang et al. [17] showed an additional Double Discrete VAE (D2-VAE) network for deep discretization and EEG latent representation learning. Using the Vector Quantized VAE (VQ-VAE), the technique explicitly provides an EEG data description codebook that can be learned. While creating a histogram of the quantisation patterns characterises the discretisation of the signal's global information, codebook queries are used to acquire the discretisation of the signal's local patterns. The repeating of single modes and blending of many modes in epileptic signals are most effectively captured by this local-global signal model. The efficacy of the suggested approach for the classification of complicated EEG signals is validated using numerous evaluation metrics and multiple epilepsy diagnostic tasks. D2-VAE significantly outperforms current techniques and has a low-dimensional yet potent quantitative representation.

Chen et al., [18] suggested a novel Hybrid Attention-enhanced Regularized Convolutional Autoencoder (HA-RCAE) feature extraction method that generates low-dimensional features, effectively preserving critical biological information while mitigating technical batch effects in EEG signals. This model is augmented with regularization techniques and a novel hybrid attention mechanism for enhanced feature extraction. Regularization techniques were employed to develop three CAE variants, and a novel HA mechanism was introduced to integrate features effectively. These features, functioning as inputs for seizure prediction and potential clinical biomarkers, exhibited robustness across datasets by mitigating batch effects and improving generalization. Coupled with classic machine learning classifiers, the proposed method outperforms advanced techniques on two epilepsy EEG datasets. This comprehensive approach offers a robust solution for seizure prediction and other clinical applications, distinguished by strong generalization and negligible batch effects, rendering it promising for clinical implementation.

Li et al. [19] developed an EEG seizure detection framework with a maximum mean discrepancy-based information maximized loss utilizing a channel-embedding spectral-temporal squeeze-and-excitation network (CE-stSENet). First, multi-level spectral and multi-scale temporal analysis are concurrently combined using CE-stSENet. A squeeze-and-excitation block variant is then used to obtain the hierarchical multi-domain depictions in an integrated way. Using features extracted from earlier subnetworks, the classification net is eventually put into practice for the identification of epileptic EEG. In particular, to address the significant overfitting issue, a maximum mean discrepancy-based data maximizing loss is synchronized via the CE-stSENet in seizure detection and the finite data distribution caused by the absence of seizure occurrences. Outcomes of omparative study on three EEG datasets compared to the standard approaches show how useful the recommended structure is in detecting epileptic EEGs. It demonstrates its powerful ability to identify seizures automatically.

Geng et al. [20] shown an effective automated seizure detection technique for intracranial EEG data using the Stockwell transform (S-transform) and bidirectional LSTM (BiLSTM) NN. The S-transform is the initial step in processing the raw EEG segments. Following that, to select and classify features, a BiLSTM receives the matrix after it has been segmented into time-frequency blocks. Postprocessing, which comprises multichannel fusion, threshold judgment, moving average filter, and collar approach, is then employed for determining classification efficiency. The proposed method is evaluated using 689 hours of intracranial EEG data from 20 individuals. The event-based evaluation provides a sensitivity of 96.30% and a false detection rate of 0.24/h. The findings show that the seizure detection method has a bright future in medical settings.

Tawhid et al. developed Convolutional LSTM (ConvLSTM), an efficient framework based on the deep spatiotemporal NN, for the diagnosis of epilepsy from EEG data [21]. Standard 19-channel EEG data are first chosen, resampled at 256 Hz, and then divided into 3-s time intervals in the suggested model. The segmented data is then input into the ConvLSTM model to differentiate between normal individuals and epileptic patients. Leave-one-out cross-validation (LOOCV) and five-fold cross-validation are used to eliminate the experiment's biases. The suggested model is qualified for usage as an automated system for epilepsy diagnosis because the experimental findings validate that it overtakes standard methods for the datasets under analysis.

3. PROPOSED METHODOLOGY

MVMD algorithm is developed for EEG signal denoising. The presence of a shared or joint frequency component in each input channel of the EEG signal is the basis for the formulation of multivariate oscillations. Attention Aggregation U-Net++ (AAUNET++) has been proposed to enhance feature extraction by incorporating a global context-aware attention mechanism. A Diffusion-Based Variational Autoencoder (DVAE) is introduced as a hybrid generative framework that unifies the probabilistic latent space learning of a VAE with the iterative noise refinement process of a DDPM. The robustness and accuracy of epileptic seizure detection are increased by this approach, which makes it easier to extract discriminative and noise-resilient EEG representations. Using the CHB-MIT database, performance metrics including ACC, PRE, Recall, and F-score were investigated in automatic seizure detection.

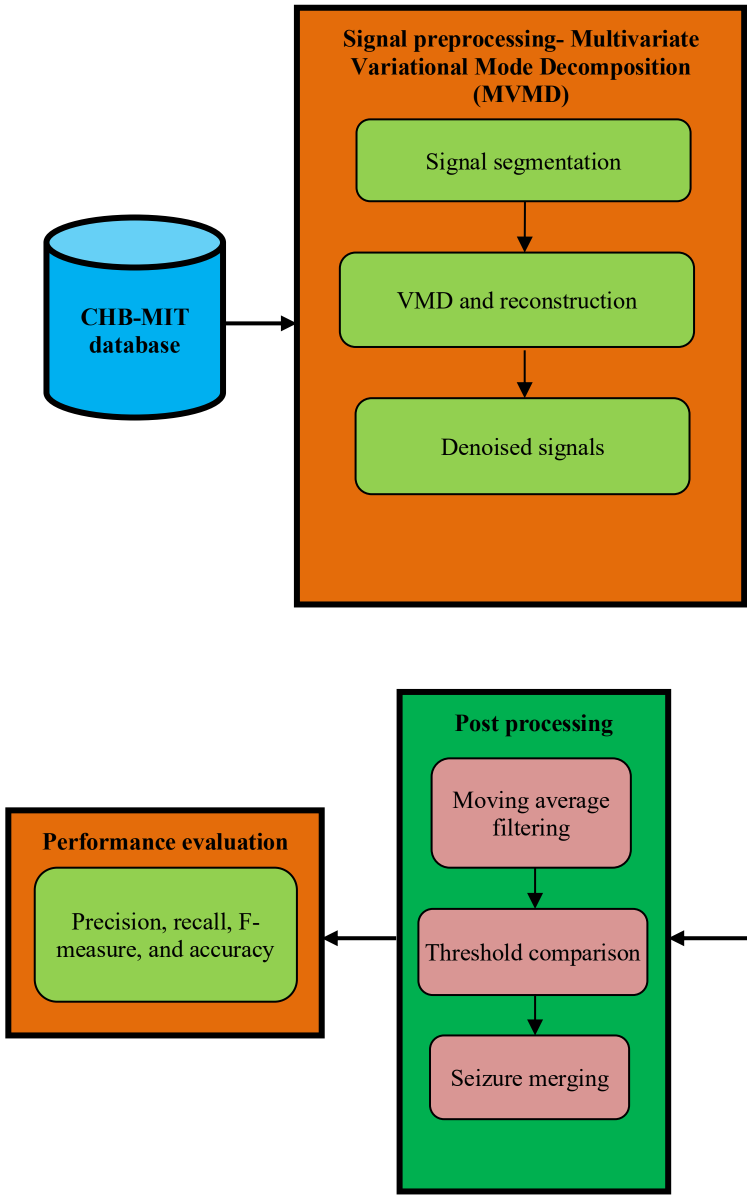

FIGURE 1. OVERALL FLOW OF SUGGESTED FRAMEWORK

3.1. Signal Pre-Processing based MVMD

For signal denoising, the variational mode decomposition (VMD) technique was created. Equation (1) [22] is used to break down an input ECG signal into a predetermined number K of intrinsic modes, $u_k(t)$.

$$x(t) = \sum_{k=1}^{K} u_k(t) \tag{1}$$

A variation model is used to select the modes that concurrently reconstructs the input ECG signal in the least squares sense and minimises the sum of frequencies of all modes $\{u_k(t)\}_{k=1}^{K}$. The unilateral frequency $u_+^k(t)$ represents ECG signals. In VMD, the frequency of every mode $u_k(t)$ is determined by harmonic mixing of ECG signal produced in the previous step with the exponential frequency ω_k, and the squared L^2 norm of the gradient

Equation (2) represents the associated constrained variational optimization problem. The decomposition is designed to provide accurate reconstruction of the ECG signal while minimising the total frequency of all modes [22].

$$\underset{\{u_k\},\{\omega_k\}}{minimize}\left\{\sum_k ||\partial_t[\,u_+^k(t)e^{-j\omega_k t}]||_2^2\right\} \tag{2}$$

Subject to $\sum_k u_k(t) = x(t)$

The partial derivative operation with regard to time is represented by the notation ∂_t; $\{u_k\}\&\{\omega_k\}$ are sets of all K modes and their center frequency. After that, an augmented Lagrangian is built to address the problem by denoising the ECG signal using equation (2). To guarantee that the constraints are precisely met in the denoising, to ensure reconstruction integrity, a quadratic term and a term with Lagrangian multipliers λ are the two penal value terms added to the Lagrangian [22]. Equation (3) represents the augmented Lagrangian function.

$$\begin{aligned}\mathscr{L}(\{\square_\square\},\{\square_\square\},\square) &= \alpha\sum_k \left\|\partial_t[u_+^k(\mathrm{t})e^{-j\omega_k t}]\right\|_2^2 \\ &+ \left\|\square(\square)-\sum_\square \square_\square(\square)]\right\|_2^2 \\ &+ \langle\lambda(t), x(t)-\sum_k u_k(t)\rangle\end{aligned} \tag{3}$$

Equation (3) can now be used to determine the Lagrangian function's saddle point, which is the original optimisation problem as stated in equation (2). Equation (3) is solved in VMD using the alternate direction method of multipliers (ADMM) methodology. A number of iterative sub-optimization tasks is used to optimize ECG signal denoising using ADMM. These sub problems are utilised to simultaneously denoised all ECG signals by iteratively minimising the cost function for individual parameters. For example, equation (4) considers the following sub-optimization problem of ECG signal denoising at the n[th] iteration in order to update $u_k(t)$.

$$\square_\square^{\square+1} \leftarrow \arg\min_{u_k}\mathscr{L}\left(\{u_{i<k}^{n+1}\},\{u_{i\geq k}^{n+1}\},\{\omega_i^n\},\lambda^n\right) \tag{4}$$

Equation (5) provides the frequency domain update seen follows once the minimisation problem of ECG signal denoising is solved in the domain in VMD (7)–(10).

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{x}(\omega)-\sum_{i\neq k}\hat{u}_i(\omega)+\dfrac{\hat{\lambda}(\omega)}{2}}{1+2\alpha(\omega-\omega_k)^2} \tag{5}$$

Equation (6) iteratively solves the following sub-optimization problem for ECG signal denoising in order to provide an update ω_k.

$$\omega_k^{n+1}(\omega) \leftarrow \arg\min_{u_k}\mathscr{L}\left(\{u_i^{n+1}\},\{\omega_{i<k}^{n+1}\},\{\omega_{i\geq k}^n\},\lambda^n\right) \tag{6}$$

Using equation (7), in the dual frequency domain, one may obtain an update for ω_k^{n+1}.

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega|\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \tag{7}$$

For multivariate ECG signals living in multi-dimensional spaces, MVMD is introduced as a generalised extension of the original VMD algorithm. C input channels are contained in an input ECG signal $x(t)$ denoted as $x(t) = [x_1(t), x_2(t),,\cdot\cdot\cdot\, x_C(t)]$. MVMD requires that the original input signal $x(t)$ be precisely reconstructed from the sum of all the extracted multivariate modes $u_k(t)$.

$$x(t) = \sum_{k=1}^{K} u_k(t) \tag{8}$$

Here, $u_k(t) = u_1(t), u_2(t),\ldots,u_C(t)$. Extracting an ensemble of multivariate modulated noises from the input ECG signal is the goal, so that: i) The total bandwidth of the modes that were extracted is the lowest; and ii) The original signal $u_k(t)$ is precisely recovered by adding the extracted modes.

The vector analytic form of $u_k(t)$ is $u_+^k(t)$. To estimate the bandwidth of $u_k(t)$, use the L_2 norm of the gradient function of the harmonically shifted $u_+^k(t)$.

Equation (9) results in the MVMD cost function f, a multidimensional extension of the VMD optimization problem cost function described in (2).

$$f = \sum_k \left\|\partial_t[e^{-j\omega_k t}u_+^k(t)]\right\|_2^2 \tag{9}$$

The whole vector $u_+^k(t)$ harmonic mixinguses only one frequency component, ω_k, according to equation (9). This results in equation (10), which provides a convenient formulation of f.

$$f = \sum_k\sum_c \left\|\partial_t[u_+^{k,c}(t)e^{-j\omega_k t}]\right\|_2^2 \tag{10}$$

Here, the ECG signal with noise eliminated for numbers c and k is indicated by $u_+^{k,c}(t)$.Unlike the vector signal in equation (10), the ECG signal with noise eliminated is $u_+^{k,c}(t)$. Equation (11) presents the constrained optimisation problem for MVMD.

$$\min_{\{u_k\},\{\omega_k\}}\left\{\sum_k\sum_c\left\|\partial_t[u_+^{k,c}(t)e^{-j\omega_k t}]\right\|_2^2\right\} \tag{11}$$

Subject to $\sum_k u_{k,c}(t) = x_C(t), c = 1\ to\ C$

It should be mentioned that the aforementioned model has many linear equality constraints, each of which relates to the total amount of input signals.

3.2. Feature extraction based attention aggregation U-net++ (AAUnet++)

Transformer U-net based on U-net++ network, using Feature Enhanced Attention (FEA) to extract the first down-sampled shallow features and the Global–Local Attention Transformer (GLAT) is based on Transformer to capture the global contextual information. The decoded features are passed through the FEA to obtain the seizure detection results. The feature mappings are produced at each stage of the TU-net presented in this study are fused with the corresponding features in the ECG signal of the decoder using a 1×1 convolution (Conv) of the 64 dimension. Another name for this technique is the leap connection. In particular, a weighted sum operation is used to combine the features produced by the residual blocks with the kind produced by the decoder GLAT [23]. The following is the expression for the weighted sum operation by equation (12),

$$F_{fusion}^{ECG} = \alpha \cdot F_{res} + (1 - \alpha) \cdot F_{GLAT} \quad (12)$$

where F_{fusion}^{ECG} is denoted as the fused features, F_{res} is denoted as the residual blocks, and F_{GLAT} is denoted as the features generated by the Transformer-based attention aggregation. The design of the feature refinement head as illustrated in Figure 2, starts with the 3×3 convolution operation on the fused features.

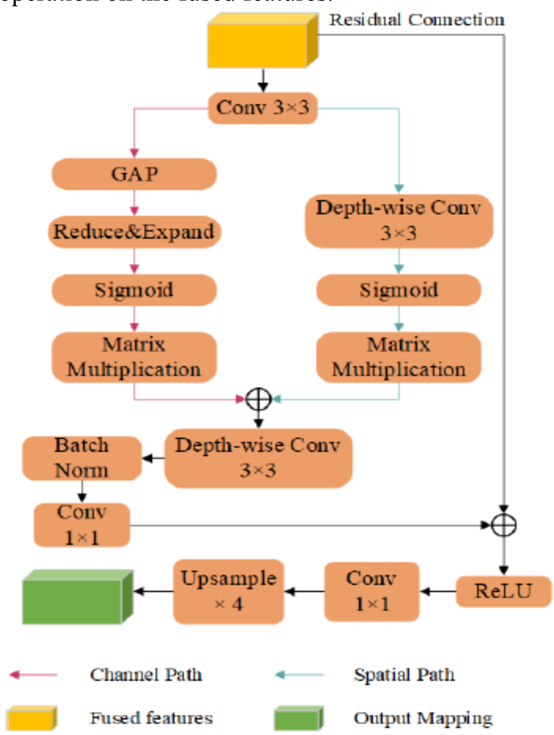

FIGURE 2. FEATURE HEAD REFINEMENT (FHR)

Construct a channel-based attention graph with a global average pooling layer in the Channel Path with channel dimensions $C \in R^{1\times1\times c}$. The scaling and cropping processes are subsequently performed by the two 1×1 Conv layers. The channel dimension is reduced by a factor of four in the 1st layer and even more in the 2nd layer. After passing the sigmoid kernel, the output features are obtained using matrix multiplication. A spatial attention map of size $C \in R^{h\times\omega\times1}$ is produced in Spatial Path using 3×3 Depth-wise Convolution, the feature map's spatial resolution is represented by $h \times \omega$. Residual connectivity is introduced to prevent the case of network degradation. After the FHR, the shallow feature extraction information can be fused with the deep global-local semantic information to improve the accuracy.

To give the embedded features multi-class and multi-scale information, transformer U-net extends the model-aware coding vector to the multi-scale application direction by adding a dynamic feature fusion block to the decoding process. Specifically, it can be set as (1) m for one vector supporting class perception and (2) n-dimensional one thermal vector supporting class perception, here, n is the ECG signal's magnification and m is a number. The equations (13-14) of the encoding can be defined as follows,

$$S_p = \begin{cases} 1, & if\ p = j, p = 1,..,n \\ 0, & else \end{cases} \quad (13)$$

$$T_k = \begin{cases} 1, & if\ k = i, k = 1,..,m \\ 0, & else \end{cases} \quad (14)$$

The class-aware vector of the ith ECG signal is denoted by T_k. The scale-aware vector at the pth scale is called S_p. A two-dimensional convolutional control layer serves as the feature fusion block to enhance the fused vectors after a flattening function flattens the three vector groups into one-dimensional vectors. Under the particular equation (15), φ is utilised as the final controller of Dynamic Head Features (DHF) to obtain the final feature head ω.

$$w = \varphi(GAP(F)\|T\|S; \theta_\varphi) \quad (15)$$

Here, the parameter count in the DHF is θ_φ, and || and Global Average Pooling (GAP) function is denoted as GAP(F), T is the temporal features, and S are combined by fusion operations. In order to accurately multi-modal features, the parameters of the feature controller based on fused features are mapped onto the dynamic head's kernel in this research. Equation (16) can be used to depict the filtering process,

$$p = \left(\left((M * we_1) * we_2\right) * we_3\right) \quad (16)$$

where the final feature head w is formed by we_1, we_2, and we_3 combined. The U-net++ feature information will significantly expand the perceptual field's range after the Dynamic Feature Fusion (DFF) module, enabling the acquisition of semantic information across a wider variety of scales. To further clarify, the use of the averaging pooling layer in obtaining the feature vector of shape helps in minimizing the spatial dimensions while maintaining the most salient data.

The GLAT is built with two parallel branches. Local semantic information is extracted using a local branch, whereas global semantic features are extracted using a global branch. Before entering the change module, the semantic information is firstly convolved by 1×1 convolution. A window operation is performed, and finally transformed into a 1D sequence, then input into the global branch. Local Branch extracts local context information using two parallel convolutional layers with convolutional kernels of sizes 1 and 3. These two convolutional layers execute extra operations prior to final summation after batch processing and normalization, and the result is a shape tensor.

Finally, the global feature information is output, and more accurate semantic information can be obtained by fusing it with the local information [24].

3.3. Classification based Diffusion based Variational Autoencoder (DVAE)

To categorize seizures caused by epilepsy, VAE is based on a simple but principled encoder-decoder design [25].

Given ECG x with a latent representation z, learning the is performed by maximizing the evidence lower bound (ELBO) on the log-likelihood log p(x).

$$\mathscr{L}(\theta,\phi) = E_{q_{\phi}(z|z)}[\log_{p_{\theta}}(x|z)] - D_{KL}[q_{\phi}(z|x)||p(z)] \tag{17}$$

Deep NN (DNN) are used to represent the approximate posterior $q_{\phi}(z|x)$ and the likelihood $(p_{\phi}(x|z))$ in the context of VAE. To train these models, the reparameterization method is crucial. The DDPM is composed of a forward detection process that progressively degrades the structure of the input ECG signal x_0 and a method $(p(x_0,T))$ for reversal epileptic seizures that learns to extract the original ECG signal x_0 from the noise input [25]. A first-order Markov chain with Gaussian transitions is used to describe the forward detection process, which is fixed during training. One can either learn or fix the schedules β_1 to β_T. Equations (18–20) provide a summary of the forward process's form.

$$q(x_i:T|x_0) = \prod_{t=1}^{r} q(x_t|x_{t-1}) \tag{18}$$

$$q(x_t|x_{t-1}) = \mathcal{N}(\sqrt{1-\beta_t}x_{t-1},\beta_t \mathrm{I}) \tag{19}$$

$$q(x_t|x_0) = \mathcal{N}(\sqrt{\bar{a}_t}x_0,(1-\bar{a}_t)\mathrm{I}) \ where\ a_t = (1-\beta_t)\ and\ \bar{a}_t = \prod_t a_t \tag{20}$$

Equations (21–22) summarise a learned Gaussian transition distribution in a first-order Markov chain can be used to parameterise the reverse process.

$$p(x_0,T) = p(x_r)\prod_{t=1}^{T} p_0(x_{t-1}|x_t) \tag{21}$$

$$p_0(x_{t-1}|x_t) = \mathcal{N}(\mu_\theta(x_t,t),\sum_{\theta}(x_t,t)) \tag{22}$$

The distribution $q(x_T|x_0)$ will approximate an isotropic Gaussian if T is sufficiently large and the variance schedule of β_t is well-behaved. Variational inference can be used to train the probabilistic system. By sampling a latent (x_0) from p(x_T) and performing the reverse procedure, a new ECG signal with epileptic seizure variations is produced from the underlying original ECG distribution.

DVAE: With an ECG signal x_0, a conditioning signal y to be modelled using a VAE, a latent representation z linked with y, and a series of T representations $x_{1:T}$ learned using a diffusion model, equation (23) can factorise the DVAE joint distribution.

$$p(x_{0,T},y,z) = p(z)p_0(y|z)p_{\phi}(x_{0:T}|y,z) \tag{23}$$

The reverse process of the conditional diffusion model is denoted as ϕ, and the VAE decoder's parameters are represented by θ.

The joint posterior $p(x_{1:T},z|y,x_0)$ is often intractable to compute. To address this, the approximate posterior is introduced. The factorized substitute posterior $q(x_{1:T},z|y,x_0)$ as defined in Equation (24):

$$q(x_{1:T},z|y,x_0) = q_{\psi}(z|y,x_0)q(x_{1:T},z|y,x_0) \tag{24}$$

The parameters of the VAE recognition network $q_{\psi}(z|y,x_0)$ is denoted as ψ. The forward process $q(x_{1:T},z|y,x_0)$ is non-trainable during training when using DDPM. Equation (25) can then be used to determine the log-likelihood of the training ECG signals.

$$\log p(x_0,y) = log\int p(x_{0,T},y,z)dx_{1:T}d_z \tag{25}$$

Optimise the ELBO that corresponds to the log-likelihood because it is challenging to estimate analytically.

4. RESULTS AND DISCUSSION

The methods for detecting epileptic seizures were developed using MatlabR2019a with an i7 CPU, 16GB RAM, and an NVIDIA RTX3050 GPU. The efficacy of the seizure detection methods was evaluated using the testing data for every case in the CHB-MIT database.

4.1. Database

The CHB-MIT scalp EEG database was acquired by Children's Hospital Boston from https://physionet.org/content/chbmit/1.0.0/. The CHB-MIT Scalp EEG Database contains long-term recordings of children with epilepsy. The most of recordings include 23 channels, and each signal is collected with 16-bit resolution at 256 Hz (with a handful having 24 or 26). The International 10–20 techniques are used to position electrodes, which ensures uniform cortical area coverage. To prepare the data for learning, each EEG channel was band-limited between 0.5 and 40 Hz using a finite impulse response (FIR) filter to minimize power-line interference, baseline drift, EMG noise, and high-frequency disturbances. Artifact-induced epochs were excluded to maintain signal quality. Following filtering, each recording was segmented into fixed-length windows of 1 second (256 samples) with 50% overlap, a configuration shown to improve temporal precision in seizure detection tasks. EEG recordings from 23 epileptics are included in the CHB-MIT database. The same individual provided Chb21 and Chb01, which were taken 1.5 years apart, and the cases were then divided into 24 groups. For every case, there are from nine to forty-two continuous EEG files accessible. Several of the files include one-hour EEG recordings, except for a few files from instances chb04, chb06, chb07, chb09, chb10, and chb23, which continue for two or four hours. Patients were assessed for a few days to describe their seizures and determine if surgery might be effective on quitting anti-seizure medication.

4.2. Performance evaluation

The following equations (26–30) can be used to calculate precision, recall/sensitivity, F-measure, and accuracy, which are utilised as assessment criteria to assess the efficiency of the seizure detection approaches.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{26}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{27}$$

$$\text{F} - \text{score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{28}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{29}$$

True Positive (TP) and True Negative (TN) are the numbers of seizure and non-seizure segments that the detection system accurately detects. False positive (FP) and false negative (FN) are EEG segments that the detection system inappropriately identified as seizures. Table 1 compares the performance of seizure detection techniques such as Bi-LSTM, Bi-GRU, CE-stSENet, Attention-based Residual Optimised Bidirectional LSTM (AROBILSTM), Stacked Multi-scale Correlation Denoising Autoencoder (SMCDAE), and DVAE in terms of aforementioned metrics.

TABLE 1. METRICS ANALYSIS OF SEIZURE DETECTION APPROACHES

Methods	Precision (%)	Recall (%)	F-Measure (%)	Accuracy (%)
Bi-LSTM	85.57	87.79	86.67	86.83
Bi-GRU	87.38	89.52	88.44	88.21
CE-stSENet	91.51	92.88	92.19	91.54
AROBILSTM	93.65	94.23	93.94	93.57
SMCDAE	94.87	95.39	95.13	96.41
DVAE	96.41	96.85	96.63	97.89

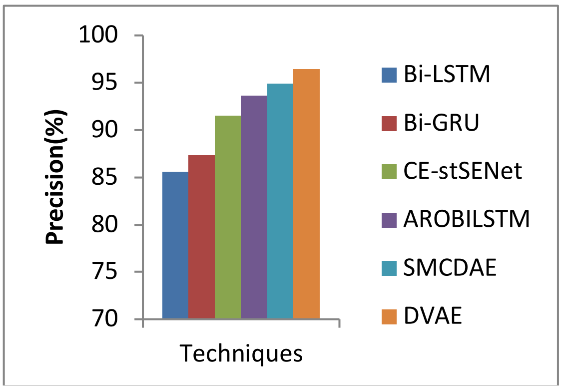

FIGURE 3. PRECISION EVALUATION OF SEIZURE DETECTION APPROACHES

The precision comparison of some standard seizure detection techniques like Bi-LSTM, Bi-GRU, CE-stSENet, AROBiLSTM, SMCDAE, and DVAE are illustrated in Figure 3. As illustrated, the proposed classifier attains the highest precision value of 96.41%, indicating a superior ability to correctly identify seizure events and lessening FP. Bi-LSTM and Bi-GRU obtain 85.57%, and 87.38%, respectively. CE-stSENet, AROBILSTM, and SMCDAE perform slightly better with a precision of 91.51%, 93.65%, and 94.87%. Proposed model has 10.84%, 9.03%, 4.90%, 2.76%, and 1.54% highest precision when compared to Bi-LSTM, Bi-GRU, CE-stSENet, AROBiLSTM, and SMCDAE.

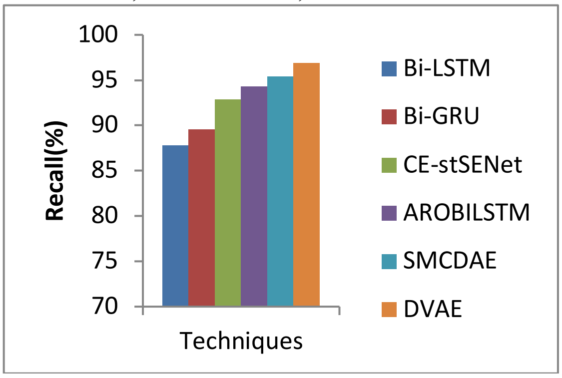

FIGURE 4. RECALL EVALUATION OF SEIZURE DETECTION APPROACHES

Recall achieved by several state-of-the-art seizure detection techniques like Bi-LSTM, Bi-GRU, CE-stSENet, AROBiLSTM, SMCDAE, and DVAE are illustrated in figure 4. As illustrated, the suggested classifier achieves the maximum recall value of 96.85%, demonstrating a higher capacity to accurately detect seizure occurrences while reducing TN and FN. Bi-LSTM and Bi-GRU obtain 87.79%, and 89.52%, respectively. CE-stSENet, AROBILSTM, and SMCDAE perform slightly better with a recall of 92.88%, 94.23%, and 95.39%. Proposed model has 9.06%, 7.33%, 3.97%, 2.62%, and 1.46% highest recall when compared to Bi-LSTM, Bi-GRU, CE-stSENet, AROBiLSTM, and SMCDAE.

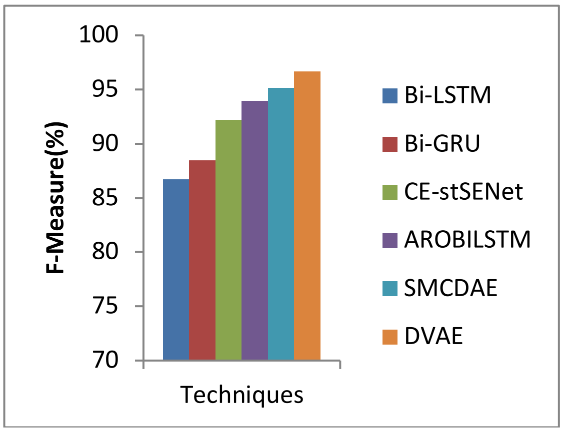

FIGURE 5. F-SCORE EVALUATION OF SEIZURE DETECTION APPROACHES

Bi-LSTM, Bi-GRU, CE-stSENet, AROBiLSTM, SMCDAE, and DVAE with respect to f-measure are illustrated in figure 5. As illustrated, the proposed classifier attains the greatest f-score of 96.63%, suggesting that recall and accuracy are harmoniously balanced. Bi-LSTM and Bi-GRU obtain 86.67% and 88.44%, respectively. CE-stSENet, AROBILSTM, and SMCDAE perform slightly better with an f-measure of 92.19%, 93.94%, and 95.13%. Proposed model has 9.96%, 8.19%, 4.44%, 2.69%, and 1.50% highest f-measure when compared to Bi-LSTM, Bi-GRU, CE-stSENet, AROBiLSTM, and SMCDAE. The proposed model has attained highest F-measure reflecting its strong ability to accurately detect seizure events with lowest false negatives.

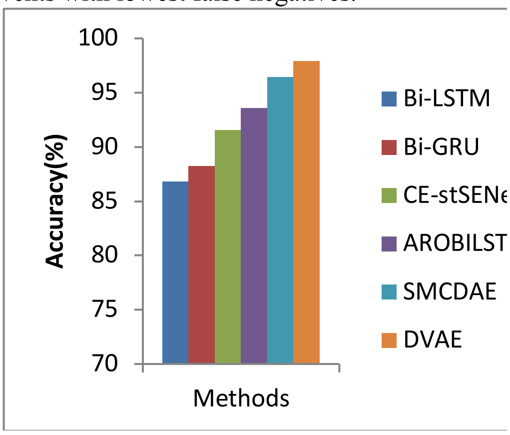

FIGURE 6. ACCURACY EVALUATION OF SEIZURE DETECTION APPROACHES

Figure 6 compares the accuracy of seizure detection techniques such as Bi-LSTM, Bi-GRU, CE-stSENet, AROBiLSTM, SMCDAE, and DVAE. Proposed classifier attains the highest accuracy of 97.89%, Bi-LSTM and Bi-GRU obtain 86.83% and 88.21%, respectively. CE-stSENet, AROBILSTM, and SMCDAE perform slightly better with accuracy of 91.54%, 93.57%, and 96.41%. Proposed model has 11.06%, 9.68%, 6.35%, 4.32%, and 1.48% highest accuracy when compared to Bi-LSTM, Bi-GRU, CE-stSENet, AROBiLSTM, and SMCDAE.

5. CONCLUSION AND FUTURE WORK

The purpose of this work is to investigate whether the CHB-MIT dataset can incorporate local EEG signals from an epilepsy center at King Abdulaziz University Hospital. The ADMM technique is used to build multivariate oscillations, and the MVMD algorithm is developed for denoising EEG signals. For multivariate ECG signals living in multi-dimensional spaces, MVMD is introduced as an extension of the VMD technique. AAUnet++ network using Feature Enhanced Attention (FEA) to extract the first down-sampled shallow features and the Global–Local Attention Transformer (GLAT) is based on Transformer to capture the global contextual information. A forward detection method and a reverse epileptic seizure mechanism that learns the DVAE and DDPM algorithms restore the original ECG signal from noisy input. The effectiveness of the seizure detection methods is assessed using precision, recall/sensitivity, F-measure, and accuracy as assessment criteria. Metrics like precision of 96.41%, recall of 96.85%, F-measure of 96.63%, and accuracy of 97.89% are used to evaluate the results. Expert systems for related illnesses involving EEG brain signals can be implemented using the suggested system. Since the suggested model is portable and the input data is tiny, it is suitable for cloud deployment and will be taken into consideration in future work. It is simple to transfer the EEG signals to the cloud for real-time processing since it can send out a warning alarm to patients and clinicians. Clinical validation and integration with therapeutic intervention systems will be part of future research.

REFERENCES

1. Alharthi, M.K., Moria, K.M., Alghazzawi, D.M. and Tayeb, H.O., 2022. Epileptic disorder detection of seizures using EEG signals. Sensors, 22(17), pp.1-18.
2. Li, R., Wang, H., Wang, L., Zhang, L., Zou, T., Wang, X., Liao, W., Zhang, Z., Lu, G. and Chen, H., 2021. Shared and distinct global signal topography disturbances in subcortical and cortical networks in human epilepsy. Human Brain Mapping, 42(2), pp.412-426.
3. Weber, D., 2025. EEG in Epilepsy. Continuum: Lifelong Learning in Neurology, 31(1), pp.38-60.
4. Islam, M.R., Zhao, X., Miao, Y., Sugano, H. and Tanaka, T., 2023. Epileptic seizure focus detection from interictal electroencephalogram: a survey. Cognitive neurodynamics, 17(1), pp.1-23.
5. Wang, C., Yuan, X. and Jing, W., 2025. Artificial intelligence in electroencephalography analysis for epilepsy diagnosis and management. Frontiers in Neurology, 16, pp.1-16.
6. Liu, W., Liu, Y., Li, S. and Chen, Y., 2023. A review of variational mode decomposition in

seismic data analysis. Surveys in Geophysics, 44(2), pp.323-355.
7. He, X., Zhou, X., Yu, W., Hou, Y. and Mechefske, C.K., 2021. Adaptive variational mode decomposition and its application to multi-fault detection using mechanical vibration signals. ISA transactions, 111, pp.360-375.
8. Li, Y., Li, X., Wang, H., Wang, S., Gu, S. and Zhang, H., 2020. Exposed aggregate detection of stilling basin slabs using attention U-net network. KSCE Journal of Civil Engineering, 24(6), pp.1740-1749.
9. Mourad, R., Diab, A., Merhi, Z., Khalil, M. and Jeannès, R.L.B., 2025. Machine and deep learning methods for epileptic seizure recognition using EEG data: A systematic review. Brain research, 1864, pp.1-48.
10. Yogarajan, G., Alsubaie, N., Rajasekaran, G., Revathi, T., Alqahtani, M.S., Abbas, M., Alshahrani, M.M. and Soufiene, B.O., 2023. EEG-based epileptic seizure detection using binary dragonfly algorithm and deep neural network. Scientific Reports, 13(1), pp.1-16.
11. Liang, S., Zhang, X., Zhao, H., Dang, Y., Hui, R. and Zhang, J., 2024. Double discrete variational autoencoder for epileptic EEG signals classification. IEEE Access, 12, pp.106567-106578.
12. Das, S., Mumu, S.A., Akhand, M.A.H., Salam, A. and Kamal, M.A.S., 2024. Epileptic seizure detection from decomposed EEG signal through 1D and 2D feature representation and convolutional neural network. Information, 15(5), pp.1-16.
13. Timothy Aboyeji, S., Wang, X., Chen, Y., Ahmad, I., Li, L., Liu, Z., Yao, C., Zhao, G., Zhang, Y., Li, G. and Chen, S., 2024. Automatic detection of epileptic seizure based on one dimensional cascaded convolutional autoencoder with adaptive window-thresholding. Journal of Neural Engineering, 21(5), pp.1-19.
14. Raghu, S., Sriraam, N., Temel, Y., Rao, S.V. and Kubben, P.L., 2020. EEG based multi-class seizure type classification using convolutional neural network and transfer learning. Neural Networks, 124, pp.202-212.
15. Dang, W., Lv, D., Rui, L., Liu, Z., Chen, G. and Gao, Z., 2021. Studying multi-frequency multilayer brain network via deep learning for EEG-based epilepsy detection. IEEE sensors journal, 21(24), pp.27651-27658.
16. Sahani, M., Rout, S.K. and Dash, P.K., 2021. Epileptic seizure recognition using reduced deep convolutional stack autoencoder and improved kernel RVFLN from EEG signals. IEEE transactions on biomedical circuits and systems, 15(3), pp.595-605.
17. Liang, S., Zhang, X., Zhao, H., Dang, Y., Hui, R. and Zhang, J., 2024. Double discrete variational autoencoder for epileptic EEG signals classification. IEEE Access, 12, pp.106567-106578.
18. Chen, Y.H., Chen, H.Y., Chen, Y.C., Chen, H.F. and Chen, H.H., 2025. Hybrid Attention-Enhanced Regularized Convolutional Autoencoder for Robust EEG Feature Extraction in Seizure Prediction. IEEE Access, vol.13, pp. 161053 – 161072.
19. Li, Y., Liu, Y., Cui, W.G., Guo, Y.Z., Huang, H. and Hu, Z.Y., 2020. Epileptic seizure detection in EEG signals using a unified temporal-spectral squeeze-and-excitation network. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28(4), pp.782-794.
20. Geng, M., Zhou, W., Liu, G., Li, C. and Zhang, Y., 2020. Epileptic seizure detection based on stockwell transform and bidirectional long short-term memory. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28(3), pp.573-580.
21. Tawhid, M.N.A., Siuly, S. and Li, T., 2022. A convolutional long short-term memory-based neural network for epilepsy detection from EEG. IEEE Transactions on Instrumentation and Measurement, 71, pp.1-11.
22. Camarena-Martinez, D., Huerta-Rosales, J. R., Amezquita-Sanchez, J. P., Granados-Lieberman, D., Olivares-Galvan, J. C., & Valtierra-Rodriguez, M. (2024). Variational Mode Decomposition-Based Processing for Detection of Short-Circuited Turns in Transformers Using Vibration Signals and Machine Learning. *Electronics*, *13*(7), pp.1-19.
23. Yang, D., Du, J., Liu, K., Sui, Y., Wang, J. and Gai, X., 2023. Construction of U-Net++ pulmonary nodule intelligent analysis model based on feature weighted aggregation. Technology and Health Care, 31(1_suppl), pp.477-486.
24. Tang, M., He, Y., Aslam, M., Akpokodje, E. and Jilani, S.F., 2025. Enhanced U-Net++ for improved semantic segmentation in landslide detection. Sensors, 25(9), pp.1-22.
25. Yu, M.-S., Jung, T.-W., Yun, D.-Y., Hwang, C.-G., Park, S.-Y., Kwon, S.-C., & Jung, K.-D. (2024). A Variational Autoencoder Cascade Generative Adversarial Network for Scalable 3D Object Generation and Reconstruction. Sensors, 24(3), pp.1-17.
26. Zhao, M., Liu, M., Ren, B., Dai, S., Sebe, N.: Denoising diffusion probabilistic models for action-conditioned 3d motion generation. In: ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 4225–4229, 2024.